Zbigniew W. Ras and Li-Shiang Tsay (Eds.)

Studies in Computational Intelligence, Volume 265

Editor-in-Chief
Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Vol. 245. Oleg Okun and Giorgio Valentini (Eds.)
Applications of Supervised and Unsupervised Ensemble Methods, 2009
ISBN 978-3-642-03998-0

Vol. 246. Thanasis Daradoumis, Santi Caballé, Joan Manuel Marquès, and Fatos Xhafa (Eds.)
Intelligent Collaborative e-Learning Systems and Applications, 2009
ISBN 978-3-642-04000-9

Vol. 247. Monica Bianchini, Marco Maggini, Franco Scarselli, and Lakhmi C. Jain (Eds.)
Innovations in Neural Information Paradigms and Applications, 2009
ISBN 978-3-642-04002-3

Vol. 248. Chee Peng Lim, Lakhmi C. Jain, and Satchidananda Dehuri (Eds.)
Innovations in Swarm Intelligence, 2009
ISBN 978-3-642-04224-9

Vol. 249. Wesam Ashour Barbakh, Ying Wu, and Colin Fyfe
Non-Standard Parameter Adaptation for Exploratory Data Analysis, 2009
ISBN 978-3-642-04004-7

Vol. 250. Raymond Chiong and Sandeep Dhakal (Eds.)
Natural Intelligence for Scheduling, Planning and Packing Problems, 2009
ISBN 978-3-642-04038-2

Vol. 251. Zbigniew W. Ras and William Ribarsky (Eds.)
Advances in Information and Intelligent Systems, 2009
ISBN 978-3-642-04140-2

Vol. 252. Ngoc Thanh Nguyen and Edward Szczerbicki (Eds.)
Intelligent Systems for Knowledge Management, 2009
ISBN 978-3-642-04169-3

Vol. 253. Roger Lee and Naohiro Ishii (Eds.)
Software Engineering Research, Management and Applications 2009, 2009
ISBN 978-3-642-05440-2

Vol. 254. Kyandoghere Kyamakya, Wolfgang A. Halang, Herwig Unger, Jean Chamberlain Chedjou, Nikolai F. Rulkov, and Zhong Li (Eds.)
Recent Advances in Nonlinear Dynamics and Synchronization, 2009
ISBN 978-3-642-04226-3

Vol. 255. Catarina Silva and Bernardete Ribeiro
Inductive Inference for Large Scale Text Classification, 2009
ISBN 978-3-642-04532-5

Vol. 256. Patricia Melin, Janusz Kacprzyk, and Witold Pedrycz (Eds.)
Bio-inspired Hybrid Intelligent Systems for Image Analysis and Pattern Recognition, 2009
ISBN 978-3-642-04515-8

Vol. 257. Oscar Castillo, Witold Pedrycz, and Janusz Kacprzyk (Eds.)
Evolutionary Design of Intelligent Systems in Modeling, Simulation and Control, 2009
ISBN 978-3-642-04513-4

Vol. 258. Leonardo Franco, David A. Elizondo, and José M. Jerez (Eds.)
Constructive Neural Networks, 2009
ISBN 978-3-642-04511-0

Vol. 259. Kasthurirangan Gopalakrishnan, Halil Ceylan, and Nii O. Attoh-Okine (Eds.)
Intelligent and Soft Computing in Infrastructure Systems Engineering, 2009
ISBN 978-3-642-04585-1

Vol. 260. Edward Szczerbicki and Ngoc Thanh Nguyen (Eds.)
Smart Information and Knowledge Management, 2009
ISBN 978-3-642-04583-7

Vol. 261. Nadia Nedjah, Leandro dos Santos Coelho, and Luiza de Macedo de Mourelle (Eds.)
Multi-Objective Swarm Intelligent Systems, 2009
ISBN 978-3-642-05164-7

Vol. 262. Jacek Koronacki, Zbigniew W. Ras, Slawomir T. Wierzchon, and Janusz Kacprzyk (Eds.)
Advances in Machine Learning I, 2009
ISBN 978-3-642-05176-0

Vol. 263. Jacek Koronacki, Zbigniew W. Ras, Slawomir T. Wierzchon, and Janusz Kacprzyk (Eds.)
Advances in Machine Learning II, 2009
ISBN 978-3-642-05178-4

Vol. 264. Olivier Sigaud and Jan Peters (Eds.)
From Motor Learning to Interaction Learning in Robots, 2009
ISBN 978-3-642-05180-7

Vol. 265. Zbigniew W. Ras and Li-Shiang Tsay (Eds.)
Advances in Intelligent Information Systems, 2009
ISBN 978-3-642-05182-1

Zbigniew W. Ras and Li-Shiang Tsay (Eds.)

Advances in Intelligent Information Systems

Springer

Zbigniew W. Ras
University of North Carolina
Department of Computer Science
Charlotte, N.C. 28223
USA
E-mail: ras@uncc.edu

Li-Shiang Tsay
NC A&T State University
Department of Electronics
Computer & Information Technology
Greensboro, NC 27411
USA
E-mail: litsay@ncat.edu

ISBN 978-3-642-05182-1 e-ISBN 978-3-642-05183-8

DOI 10.1007/978-3-642-05183-8

Studies in Computational Intelligence ISSN 1860-949X

Library of Congress Control Number: 2009940422

Typeset & *Cover Design:* Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed in acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

Intelligent Information Systems (IIS) can be defined as the next generation of Information Systems (IS) developed as a result of integration of AI and database (DB) technologies. IIS embody knowledge that allows them to exhibit intelligent behavior, allows them to cooperate with users and other systems in problem solving, discovery, retrieval, and manipulation of data and knowledge. For any IIS to serve its purpose, the information must be available when it is needed. This means that the computing systems used to store data and process the information, and the security controls used to protect it must be functioning correctly. This book covers some of the above topics and it is divided into four sections: Classification, Approximation and Data Security, Knowledge Management, and Application of IIS to medical and music domains.

The first section contains three contributions in the area of Classification.

- Its first chapter is written by F. Rioult, B. Zanuttini, and B. Cremilleux. It studies the impact of generalizing the syntax of association rules used in classification processes. Authors investigate rules with disjunctive conclusions, which can be interpreted as (i) rules prescribing a class for an object but with the absence of some attributes as premises, or (ii) rules prescribing the negation of a class. Authors also investigate well-founded definitions of such rules, generalizing the studies on classical rules (about minimal premises, maximal conclusions, etc.), as well as their theoretical properties and their extraction from datasets.
- Many algorithms of machine learning use an entropy measure as optimization criterion. In some real world applications, the use of such entropy measures without precautions, could lead to inconsistent results. Indeed, the measures of entropy are built upon some assumptions which are not fulfilled in many real cases. For instance, the misclassification costs are assumed to be the same for all classes and also, the worst case (maximum of the entropy) is obtained for the uniform distribution. In the second chapter of this book, authors show that classical entropy measures

are not suitable for building a predictive model. Then, they examine the main axioms that define an entropy and discuss their inadequacy in machine learning. This leads them to propose a new entropy measures that possess more suitable proprieties.

- In the third chapter titled “Algebra and Topology for Dominance-based Rough Set Approach (DRSA)", authors present algebraic representations of DRSA in terms of generalizations of several algebras already used to represent the classical rough set approach, namely: bipolar de Morgan Brouwer-Zadeh distributive lattice, bipolar Nelson algebra, bipolar Heyting algebra, bipolar double Stone algebra, bipolar three-valued Lukasiewicz algebra, and bipolar Wajsberg algebra. DRSA was first introduced as a generalization of the rough set approach for dealing with multiple criteria decision analysis. In this chapter, authors also present an algebraic model for ordinal classification.

The second part of the book contains four contributions in the areas of Approxi mation & Data Security.

- In the first chapter authors propose a method to tackle the complexity of the induction process in the setting of multi-relational learning. In particular, an approximate reasoning technique and an approximate query evaluation have been used to keep low the complexity of relational learning and to tackle the NP-completeness of the deduction process. The implementation of those approximate reasoning techniques has been applied to solve both the induction and clustering problem on real-world datasets.
- In the second chapter author introduces a novel approximate query answering technique for OLAP that is based on an innovative analytical interpretation of multidimensional data cubes, and the use of the well-known Least Squares Approximation (LSA) method in order to build the so-called analytical synopsis data structure Δ-Syn. The benefits deriving from adopting Δ-Syn within the core layer of modern OLAP server platforms are confirmed by a comprehensive experimental evaluation of the performance of Δ-Syn on both synthetic and real-life data cubes that clearly show the superiority of Δ-Syn in comparison with state-of-the-art approximate query answering techniques like histograms, wavelets and random sampling.
- In the third chapter, written by F. Massacci, J. Mylopoulos, and N. Zannone, authors present a comprehensive and updated description of SI* modeling language to deal with security and trust, and the Secure Tropos methodology for designing secure software systems. SI* and Secure Tropos have been adopted in several European, national, and local projects for the analysis of early organizational and security requirements in different application domains.
- The fourth chapter, written by M. Baig, J. Li, J. Liu, H. Wang, and J. Wang , concerns human genomic data which may be maliciously used to find out the genetic tendency for a disease, and even to track descendents

and relatives of the individual. Authors review some techniques for protecting privacy in sharing human genomic data and discuss related problems and challenges.

The third part of the book contains four contributions in the area of Knowledge Management.

- In the first chapter, C. Meghini and N. Spyratos show that understanding collections in Digital Libraries (DL) as abstraction mechanisms endowed with intension and extension leads not only to a simpler, conceptually well founded DL model, but also to a more powerful one, which can still be efficiently managed.
- The second chapter, written by A.C. Acar and A. Motro, addresses the problem of interpreting distributed sets of queries, by using the well-researched architecture of virtual databases. Virtual database architecture integrates a set of local databases by means of a global database scheme which is mapped into the local databases. A query submitted to the virtual database (based on the global scheme) is decomposed into queries against the local databases, and the corresponding answers are assembled in an answer to the original query. The entire process is transparent to the user.
- The next chapter, written by C.J. Butz, W.Yan, P. Lingras, and Y.Y. Yao, refines the presentation of variable elimination in discrete Bayesian networks by showing that the probability distributions constructed during this process are a special case of the distributions stated in the literature.
- The fourth chapter is written by A. Tzacheva and it presents an improved method for reducing the space of action rules under the assumption that attributes in a decision system are hierarchical. The agglomerative clustering method generates summaries which are new compact action rules representing clusters of syntactically similar rules.

The last part of this book contains four contributions in the application areas: medicine and music.

- The first chapter of this section presents SIGHT, an implemented system for providing individuals who have sight impairments with access to bar charts that appear in popular media such as magazines and newspapers. Rather than attempting to translate bar charts into an alternative medium such as sound or touch, the system attempts to provide the user with the high-level knowledge that is conveyed. SIGHT is implemented as a Browser Helper Object for Internet Explorer. Using JAWS screen reading software, a web page is read to the user and, when a bar chart is encountered, a Bayesian reasoning system is activated that identifies the graphic's intended message which is then conveyed to the user via speech.
- In the second chapter authors present conceptual development of an ontology for mental health. The ontology is intended to support a comprehensive clinical decision support system (CDSS) for psychiatric rehabilitation. Protege 3.4 was used as the ontology editor in their

development and RacerPro was used to discover and correct the inconsistencies during development process. A rule-based inference engine is used in the ontology to arrive at clinical judgments and treatment recommendations. The ontology and the rule-base were tested with sample patient information for various conditions relevant to psychiatric rehabilitation, like schizophrenia, affective disorders and borderline personality disorder. The diagnosis generated by their system was verified to be correct based on the patient conditions. The system was also tested when patient has other medical conditions like stroke, pregnancy etc. along with psychiatric conditions. The system was also tested when a patient has some psychiatric symptoms along with physical conditions that could lead to psychiatric symptoms.

- In the third chapter, based on the experimental results, authors propose a new framework for music information retrieval with multiple classifiers trained on different features. Inspired by human recognition experience, timbre estimation based on hierarchical structure of musical instrument families is investigated. A framework for timbre automatic indexing based on cascade classification system is also proposed.
- Many papers in music information retrieval domain present results of automatic classification of instruments playing in a given sound piece, some of them compare classification types, some show results obtained on different training sets. In the third chapter, authors show how to develop new temporal features and next they check if by adding them to existing features used by music information retrieval system, called MIRAI, they improve its classifiers for automatics indexing of polyphonic music by instruments.

We wish to express our thanks to all the authors who contributed the above fifteen chapters to this book.

August 2009

Z.W. Raś
L.-S. Tsay

Contents

Part III: Knowledge Management

Part IV: Applications